

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> :		A1	(11) International Publication Number: WO 98/23743
C12N 15/12, C07K 14/47, 16/18, A61K 38/17, C12Q 1/68, G01N 33/68			(43) International Publication Date: 4 June 1998 (04.06.98)
(21) International Application Number: PCT/GB97/03302		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 1 December 1997 (01.12.97)			
(30) Priority Data: 9624905.7 29 November 1996 (29.11.96) GB			
(71) Applicant (for all designated States except US): MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB).			
(72) Inventors; and		Published	
(75) Inventors/Applicants (for US only): DAVIES, Kay, Elizabeth [GB/GB]; University of Oxford, Genetics Laboratory, Dept. of Biochemistry, South Parks Road, Oxford OX1 3QU (GB). THEODOSIOU, Aspasia [GR/GB]; University of Oxford, Genetics Laboratory, Dept. of Biochemistry, South Parks Road, Oxford OX1 3QU (GB).		With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.	
(74) Agent: WOODS, Geoffrey, Corlett; J.A. Kemp & Co., 14 South Square, Gray's Inn, London WC1R 5LX (GB).			
(54) Title: MURINE GUANINE NUCLEOTIDE EXCHANGE FACTOR - (MNGEF) AND HUMAN HOMOLOGUES THEREOF			
(57) Abstract			
The present invention concerns the cloning of a murine guanine nucleotide exchange factor designated MNGEF and a human homologue thereof. Polynucleotide probes derived from the nucleotide sequence of MNGEF and antibodies that recognise MNGEF are also provided.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Republic of Korea	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

- 1 -

MURINE GUANINE NUCLEOTIDE EXCHANGE FACTOR - (MNGEF)  
AND HUMAN HOMOLOGUES THEREOF

Field of the Invention

The present invention relates to MNGEF, a member of the family of regulators of small GTP-binding proteins, and homologues of MNGEF.

Background to the invention

The superfamily of low molecular mass GTP-binding proteins (also known as G proteins), for which ras proteins are prototypes, has been implicated in the regulation of diverse biological activities. In addition to their involvement in regulating many aspects of growth and differentiation, members of this superfamily play an important role in the control of the cytoskeleton and in the regulation of protein trafficking between various membrane-bound compartments in the cell.

These proteins function as binary switches, being 'on' in the GTP-bound state and 'off' in the GDP-bound state. Cycling between these two forms is controlled by various accessory proteins. The guanine nucleotide exchange factors (GEFs), promote the exchange of GDP for GTP, thus activating the proteins whereas, the GTPase-activating proteins (GAPS) and GDP-dissociation inhibitory factors (GDIs) are negative modulators. The Ras-like proteins are divided into six main families, based on their sequences: Rab, Arf, Sar, Ran, Rho and Ras.

Until recently, the Rho GTPases (such as Rac, Rho, Cdc42) were thought to be primarily involved in the organisation of the actin cytoskeleton. However, it has become evident that they play a critical role in controlling cell proliferation and progress has been made in identifying signalling cascades involving the Rho family members.

A family of cell growth regulatory proteins and oncogene products have been discovered for which the Dbl oncprotein is a prototype (Eva and Aaronson (1985) *Nature* 316, 273-275). These proteins are putative guanine nucleotide exchange factors for the Rho GTPases. They all contain a Dbl homology domain (DH) in tandem with a pleckstrin homology domain (PH), and seem to activate specific members of the Rho family to elicit a variety of biological functions in the cell. The DH domain is responsible for binding and activating the G proteins thus mediating downstream signalling events, whereas the PH

- 2 -

domain is thought to play a role in targeting these guanine nucleotide exchange factors to specific cellular locations in order to carry out the signalling function.

Since the initial identification of Dbl as a GEF for Rho GTPases, an increasing number of oncogene products and growth regulatory molecules have been shown to contain those two domains in tandem. Many of them, such as Bcr which is involved in the chromosomal rearrangements in chronic myelogenous leukaemia, Cdc24, Ras guanine nucleotide release factor and Vav have been implicated in cell growth regulation. Others, including Ect-2, Tim, Ost and Lbc were discovered, by virtue of their transforming capability, through gene transfer methods.

10

#### Disclosure of the Invention

Here, we report the isolation and preliminary characterisation of 3 overlapping mouse cDNAs (designated MNGEF1, MNGEF2 and MNGEF3), which show homology to the TIM gene (Transforming Immortalized Mammary, Chan *et al.*, (1994) *Oncogene* 9, 1057-1063) of the family of regulators of small GTP-binding proteins. The homology is observed at both the amino acid and nucleotide levels. However, the size of the transcript observed by Northern analysis and the expression pattern of MNGEF2 is markedly different to that of TIM, suggesting that this is a novel, neuronal-specific member of the above family of genes. In addition, MNGEF1 and MNGEF2 contain a trinucleotide repeat. Together with the high expression pattern of MNGEF2 in brain, the presence of the triplet repeat and the homology to TIM, these cDNAs present potential candidates for disease related genes.

We also report the cloning and sequencing of a fragment of the human homologue of MNGEF. Substantial homology is observed at both the amino acid and nucleotide levels between murine MNGEF and its human homologue NGEF.

25

The MNGEF3 clone is 1.35 kb and is contained completely within the MNGEF1 cDNA which is 2.3 kb. MNGEF2 is the longest clone (2.8kb) but contains a 92bp unspliced intron within it (from nucleotides 1816 to 1907 of SEQ. ID No. 3), resulting in a premature termination codon. MNGEF1 does not contain this intron and therefore its ORF extends beyond the stop codon of MNGEF2. From the sequences of MNGEF1 and MNGEF2 we conclude that the cDNA designated MNGEF consists of 2741 bp (2833 bp minus 92 bp) which results in an ORF of 554 amino acids.

- 3 -

The murine MNGEF cDNA sequence is set out as SEQ. ID No. 1. The amino acid sequence of the ORF from nucleotides 343 to 2004 is set out as SEQ. ID No. 2. The murine MNGEF2 cDNA sequence, which includes the 92 bp intron, is set out as SEQ ID No. 3. The amino acid sequence of the ORF from nucleotides 343 to 1860 is set out as SEQ. ID No. 4.

5. The murine MNGEF1 cDNA sequence is set out as SEQ. ID No. 5. The amino acid sequence of the ORF from nucleotides 2 to 1609 is set out as SEQ. ID No. 6. The partial human NGEF cDNA sequence is set out as SEQ. ID No. 7. The amino acid sequence of the ORF from nucleotides 3 to 803 is set out as SEQ ID No. 8.

Thus the invention provides a murine guanine nucleotide exchange factor designated  
10 MNGEF, a human homologue thereof designated human NGEF or other mammalian homologue thereof which guanine nucleotide exchange factor is encoded by a cDNA sequence obtainable from a mammalian brain cDNA library, said DNA sequence being selectively detectable with a murine DNA sequence as shown in SEQ ID Nos. 1, 3 or 5 or a human DNA sequence as shown in SEQ ID No. 7.

15 The protein preferably has one or more of the additional features:

- (1) it comprises a Dbl homology domain having substantial homology to amino acids 124 to 306 of SEQ ID No. 2;
- (2) it comprises a pleckstrin homology domain having substantial homology to amino acids 333 to 445 of SEQ ID No. 2;
- 20 (3) it comprises an SH3 domain (Src homology 3 domain) having substantial homology to amino acids 456 to 517 of SEQ ID No. 2
- (4) it is found predominantly in neuronal cell types;
- (5) it is encoded by an mRNA of approximately 2.7 kb;
- (6) it promotes the exchange of GDP for GTP by low molecular mass GTP-
- 25 binding proteins; and
- (7) it comprises a polyglutamine region.

The term "selectively detectable" means that the cDNA used as a probe is used under conditions where a target cDNA of the invention is found to hybridize to the probe at 30 a level significantly above background. The background hybridization may occur because of other cDNAs present in the brain cDNA library. In this event background implies a level of signal generated by interaction between the probe and a non-specific cDNA member of

- 4 -

the library which is less than 10 fold, preferably less than 100 fold as intense as the specific interaction observed with the target cDNA. The intensity of interaction may be measured, for example, by radiolabelling the probe, e.g. with  $^{32}\text{P}$ . Suitable conditions may be found by reference to the Examples.

5 Accordingly, in a first aspect, the invention provides the MNGEF protein of SEQ ID. 2, 4, 6 and homologues thereof, polypeptide fragments thereof, as well as antibodies capable of binding the MNGEF protein or polypeptide fragments thereof. The invention also provides the human NGEF protein of SEQ. ID. No. 8 and homologues thereof, polypeptide fragments thereof, as well as antibodies capable of binding the human NGEF 10 protein or polypeptide fragments thereof. Human NGEF proteins, homologues and fragments thereof, are also included in references below to polypeptides of the invention.

In another aspect, the present invention provides a polynucleotide in substantially 15 isolated form capable of hybridising selectively to any one of SEQ ID Nos. 1, 3, 5 or 7 or to the complement (i.e. opposite strand) thereof. The present invention also provides a polynucleotide in substantially isolated form capable of hybridising selectively to any one of SEQ ID Nos. 1, 3, 5 or 7 or to the complement (i.e. opposite strand) thereof. Also provided are polynucleotides encoding polypeptides of the invention. Such polynucleotides will be referred to as a polynucleotide of the invention. A polynucleotide of the invention includes DNA of SEQ ID Nos. 1, 3, 5 and fragments thereof capable of selectively hybridising to the 20 gene encoding MNGEF. A polynucleotide of the invention also includes DNA of SEQ ID No 7 and fragments thereof capable of selectively hybridising to the gene encoding human NGEF.

In a further aspect, the invention provides recombinant vectors carrying a 25 polynucleotide of the invention, including expression vectors, and methods of growing such vectors in a suitable host cell, for example under conditions in which expression of a protein or polypeptide encoded by a sequence of the invention occurs.

In an additional aspect, the invention provides kits comprising polynucleotides, 30 polypeptides or antibodies of the invention and methods of using such kits in diagnosing the presence of absence of MNGEF, human NGEF and their homologues, or variants thereof, including deleterious MNGEF and human NGEF mutants.

Detailed description of the invention.A. Polynucleotides.

In the following description, it should be understood that references to MNGEF refer 5 additionally to MNGEF1, MNGEF2, MNGEF3 and human NGEF. Polynucleotides of the invention may comprise DNA or RNA. They may be single or double stranded. They may also be polynucleotides which include within them synthetic or modified nucleotides. A number of different types of modification to oligonucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones, addition of acridine or 10 polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the polynucleotides described herein may be modified by any method available in the art. Such modifications may be carried out in order to enhance the *in vivo* activity or life span of polynucleotides of the invention.

Polynucleotides of the invention capable of selectively hybridising to the DNA of 15 SEQ ID No. 1 will be generally at least 70%, preferably at least 80 or 90% and more preferably at least 95% homologous to the corresponding DNA of SEQ ID No. 1 over a region of at least 20, preferably at least 25 or 30, for instance at least 40, 60 or 100 or more contiguous nucleotides. Preferred polynucleotides of the invention will comprise regions homologous to the DH domain of MNGEF, from nucleotides 712 to 1260 of SEQ ID No. 1, 20 preferably at least 80 or 90% and more preferably at least 95% homologous to the DH domain of MNGEF. Preferred polynucleotides of the invention will also comprise regions homologous to the PH domain of MNGEF, from nucleotides 1339 to 1677 of SEQ ID No. 1, preferably at least 80 or 90% and more preferably at least 95% homologous to the PH domain of MNGEF. Preferred polynucleotides of the invention will further comprise 25 regions homologous to the SH3 domain of MNGEF, from nucleotides 1708 to 1893 of SEQ ID No 1, preferably at least 80 or 90% and more preferably at least 95% homologous to the SH3 domain of MNGEF.

It is to be understood that skilled persons may, using routine techniques, make 30 nucleotide substitutions that do not affect the polypeptide sequence encoded by the polynucleotides of the invention to reflect the codon usage of any particular host organism in which the polypeptides of the invention are to be expressed.

- 6 -

Any combination of the above mentioned degrees of homology and minimum sizes may be used to define polynucleotides of the invention, with the more stringent combinations (i.e. higher homology over longer lengths) being preferred. Thus for example a polynucleotide which is at least 80% homologous over 25, preferably 30 nucleotides 5 forms one aspect of the invention, as does a polynucleotide which is at least 90% homologous over 40 nucleotides.

Polynucleotides of the invention may be used to produce a primer, e.g. a PCR primer, a primer for an alternative amplification reaction, a probe e.g. labelled with a revealing label by conventional means using radioactive or non-radioactive labels, or the 10 polynucleotides may be cloned into vectors. Such primers, probes and other fragments will be at least 15, preferably at least 20, for example at least 25, 30 or 40 nucleotides in length, and are also encompassed by the term polynucleotides of the invention as used herein.

Polynucleotides such as a DNA polynucleotide and primers according to the invention may be produced recombinantly, synthetically, or by any means available to those 15 of skill in the art. They may also be cloned by standard techniques.

In general, primers will be produced by synthetic means, involving a step wise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art.

Longer polynucleotides will generally be produced using recombinant means, for 20 example using a PCR (polymerase chain reaction) cloning techniques. This will involve making a pair of primers (e.g. of about 15-30 nucleotides) to a region of the MNGEF gene which it is desired to clone, bringing the primers into contact with mRNA or cDNA obtained from an animal or human cell (e.g. a brain cell), performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating 25 the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme recognition sites so that the amplified DNA can be cloned into a suitable cloning vector.

Such techniques may be used to obtain all or part of the MNGEF sequence described 30 herein. Genomic clones containing the MNGEF gene and its introns and promoter regions may also be obtained in an analogous manner, starting with genomic DNA from an animal

- 7 -

or human cell, e.g. a brain cell.

Although in general the techniques mentioned herein are well known in the art, reference may be made in particular to Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (1989) and Ausubel *et al.*, Current Protocols in Molecular Biology (1995), John Wiley & Sons, Inc.

Polynucleotides which are not 100% homologous to the sequences of the present invention but fall within the scope of the invention can be obtained in a number of ways. Other murine allelic variants of the MNGEF sequence described herein may be obtained for example by probing genomic DNA libraries made from a range of individuals, for example 10 individuals from different populations. In addition, other animal, particularly mammalian (e.g. rat or rabbit, more particularly primate), homologues of MNGEF may be obtained and such homologues and fragments thereof in general will be capable of selectively hybridising to SEQ ID No. 1. Such sequences may be obtained by probing cDNA libraries made from dividing cells or tissues or genomic DNA libraries from other animal species, and probing 15 such libraries with probes comprising all or part of SEQ ID. 1 under conditions of medium to high stringency (for example 0.03M sodium chloride and 0.03M sodium citrate at from about 50°C to about 60°C). Nucleic acid probes comprising all or part of SEQ ID No. 7 may be used to probe cDNA libraries from primate species, preferably humans, to obtain homologues of MNGEF. In particular nucleic acid probes comprising all or part of SEQ ID 20 No. 7 may be used to probe cDNA libraries from humans, to obtain the full-length cDNA encoding human NGEF or a homologue thereof.

Allelic variants and species homologues may also be obtained using degenerate PCR which will use primers designed to target sequences within the variants and homologues encoding conserved amino acid sequences. Conserved sequences can be predicted from 25 aligning the MNGEF amino acid sequence with that of TIM. The primers will contain one or more degenerate positions and will be used at stringency conditions lower than those used for cloning sequences with single sequence primers against known sequences. In particular, primers can be designed to target the DH, PH and SH3 domains described above.

Alternatively, such polynucleotides may be obtained by site directed mutagenesis of 30 the MNGEF sequences or allelic variants thereof. This may be useful where for example silent codon changes are required to sequences to optimise codon preferences for a

- 8 -

particular host cell in which the polynucleotide sequences are being expressed. Other sequence changes may be desired in order to introduce restriction enzyme recognition sites, or to alter the property or function of the polypeptides encoded by the polynucleotides. Further changes may be desirable to represent particular coding changes found in MNGEF which give rise to mutant MNGEF genes which have lost their regulatory function. Probes based on such changes can be used as diagnostic probes to detect such MNGEF mutants.

The invention further provides double stranded polynucleotides comprising a polynucleotide of the invention and its complement.

Polynucleotides or primers of the invention may carry a revealing label. Suitable labels include radioisotopes such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , enzyme labels, or other protein labels such as biotin. Such labels may be added to polynucleotides or primers of the invention and may be detected using techniques known per se.

Polynucleotides or primers of the invention or fragments thereof labelled or unlabelled may be used by a person skilled in the art in nucleic acid-based tests for detecting or sequencing MNGEF and its homologues in the human or animal body.

Such tests for detecting generally comprise bringing a biological sample containing DNA or RNA into contact with a probe comprising a polynucleotide or primer of the invention under hybridising conditions and detecting any duplex formed between the probe and nucleic acid in the sample. Such detection may be achieved using techniques such as PCR or by immobilising the probe on a solid support, removing nucleic acid in the sample which is not hybridised to the probe, and then detecting nucleic acid which has hybridised to the probe. Alternatively, the sample nucleic acid may be immobilised on a solid support, and the amount of probe bound to such a support can be detected. Suitable assay methods of this any other formats can be found in for example WO89/03891 and WO90/13667.

Tests for sequencing MNGEF and its homologues include bringing a biological sample containing target DNA or RNA into contact with a probe comprising a polynucleotide or primer of the invention under hybridising conditions and determining the sequence by, for example the Sanger dideoxy chain termination method (see Sambrook *et al.*).

Such a method generally comprises elongating, in the presence of suitable reagents, the primer by synthesis of a strand complementary to the target DNA or RNA and

- 9 -

selectively terminating the elongation reaction at one or more of an A, C, G or T/U residue; allowing strand elongation and termination reaction to occur; separating out according to size the elongated products to determine the sequence of the nucleotides at which selective termination has occurred. Suitable reagents include a DNA polymerase enzyme, the 5 deoxynucleotides dATP, dCTP, dGTP and dTTP, a buffer and ATP. Dideoxynucleotides are used for selective termination.

Tests for detecting or sequencing MNGEF, or its homologue, in a biological sample may be used to determine MNGEF sequences within cells in individuals who have, or are suspected to have, an altered MNGEF gene sequence, for example within cancer cells 10 including leukaemia cells and solid tumours such as breast, ovary, lung, colon, pancreas, testes, liver, brain, muscle and bone tumours or within cells from the nervous system of individuals suffering from neurological disorders.

In addition, the discovery of MNGEF will allow the role of this gene in hereditary diseases to be investigated. In general, this will involve establishing the status of MNGEF, 15 or its homologue (e.g. using PCR sequence analysis), in cells derived from animals or humans with, for example, neurological disorders or neoplasms.

The probes of the invention may conveniently be packaged in the form of a test kit in a suitable container. In such kits the probe may be bound to a solid support where the assay format for which the kit is designed requires such binding. The kit may also contain 20 suitable reagents for treating the sample to be probed, hybridising the probe to nucleic acid in the sample, control reagents, instructions, and the like.

The present invention also provides polynucleotides encoding the polypeptides of the invention described below. Because such polynucleotides will be useful as sequences for recombinant production of polypeptides of the invention, it is not necessary for them to be selectively hybridisable to the sequence of any one of SEQ ID Nos. 1, 3, 5 or 7 although 25 this will generally be desirable. Otherwise, such polynucleotides may be labelled, used, and made as described above if desired. Polypeptides of the invention are described below.

#### B. Polypeptides.

30 Polypeptides of the invention include polypeptides in substantially isolated form which comprise the sequence set out in SEQ ID Nos. 2, 4, 6 or 8. Polypeptides further

- 10 -

include variants of such sequences, including naturally occurring allelic variants and synthetic variants which are substantially homologous to said polypeptides. In this context, substantial homology is regarded as a sequence which has at least 70%, e.g. 80% or 90% amino acid homology (identity) over 30 amino acids with the sequence of SEQ ID No. 2.

5 Polypeptides also include other those encoding MNGEF homologues, and variants thereof as defined above, from other species including animals such as mammals (e.g. mice, rats or rabbits), especially primates, more especially humans. MNGEF homologues include human NGEF.

10 Polypeptides of the invention also include fragments of the above mentioned full length polypeptides and variants thereof, including fragments of the sequences set out in SEQ ID Nos. 2, 4, 6 or 8. Preferred fragments include those which include an epitope. Suitable fragments will be at least about 5, e.g. 10, 12, 15 or 20 amino acids in size. Polypeptide fragments of the MNGEF and human NGEF proteins and allelic and species variants thereof may contain one or more (e.g. 2, 3, 5, or 10) substitutions, deletions or 15 insertions, including conserved substitutions.

Conserved substitutions may be made according to the following table which indicates conservative substitutions, where amino acids on the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
	Polar - charged	D E
		K R
AROMATIC		H F W Y
OTHER		N Q D E

- 11 -

Epitopes may be determined, for example, by techniques such as peptide scanning techniques as described by Geysen *et al*, 1986.

Polypeptides of the invention may be in a substantially isolated form. It will be understood that the polypeptide may be mixed with carriers or diluents which will not interfere with the intended purpose of the polypeptide and still be regarded as substantially isolated. A polypeptide of the invention may also be in a substantially purified form, in which case it will generally comprise the polypeptide in a preparation in which more than 90%, e.g. 95%, 98% or 99% of the polypeptide in the preparation is a polypeptide of the invention. Polypeptides of the invention may be modified for example by the addition of histidine residues to assist their purification or by the addition of a signal sequence to promote their secretion from a cell.

A polypeptide of the invention may be labelled with a revealing label. The revealing label may be any suitable label which allows the polypeptide to be detected. Suitable labels include radioisotopes, e.g. <sup>125</sup>I, enzymes, antibodies, polynucleotides and linkers such as biotin. Labelled polypeptides of the invention may be used in diagnostic procedures such as immunoassays in order to determine the amount of a polypeptide of the invention in a sample. Polypeptides or labelled polypeptides of the invention may also be used in serological or cell mediated immune assays for the detection of immune reactivity to said polypeptides in animals and humans using standard protocols.

A polypeptide or labelled polypeptide of the invention or fragment thereof may also be fixed to a solid phase, for example the surface of an immunoassay well or dipstick. Such labelled and/or immobilised polypeptides may be packaged into kits in a suitable container along with suitable reagents, controls, instructions and the like. Such polypeptides and kits may be used in methods of detection of antibodies to the MNGEF or human NGEF proteins or their allelic or species variants by immunoassay.

Immunoassay methods are well known in the art and will generally comprise:

- (a) providing a polypeptide comprising an epitope bindable by an antibody against said protein;
- (b) incubating a biological sample with said polypeptide under conditions which allow for the formation of an antibody-antigen complex; and
- (c) determining whether antibody-antigen complex comprising said polypeptide

- 12 -

is formed.

Polypeptides of the invention may be made by synthetic means (e.g. as described by Geysen *et al.*, 1996) or recombinantly, as described below.

Particularly preferred polypeptides of the invention include those spanning or within the DH, PH or DH3 homology domains or sequences substantially homologous thereto. Preferred polypeptides comprise regions showing substantial homology to the DH domain of MNGEF represented as amino acids 124 to 306 of SEQ ID No. 2. Preferred polypeptides will also comprise regions showing substantial homology to the PH domain of MNGEF represented as amino acids 333 to 445 of SEQ ID No. 2. Preferred polypeptides will further comprise regions showing substantial homology to the SH3 domain of MNGEF represented as amino acids 456 to 517 of SEQ ID No. 2. Fragments as defined above from this region are particularly preferred. The polypeptides and fragments thereof may contain amino acid alterations as defined above.

Polypeptides of the invention may be used in *in vitro* or *in vivo* cell culture systems to study the role of MNGEF, human NGEF and their homologues in disease. For example, truncated or modified MNGEF may be introduced into a cell to disrupt the normal functions which occur in the cell. The polypeptides of the invention may be introduced into the cell by *in situ* expression of the polypeptide from a recombinant expression vector (see below). The expression vector optionally carries an inducible promoter to control the expression of the polypeptide.

The use of mammalian host cells is expected to provide for such post-translational modifications (e.g. myristylation, glycosylation, truncation, ligation and tyrosine, serine or threonine phosphorylation) as may be needed to confer optimal biological activity on recombinant expression products of the invention. Such cell culture systems in which polypeptide of the invention are expressed may be used in assay systems to identify candidate substances which interfere with or enhance the functions of the polypeptides of the invention in the cell.

#### C. Vectors.

Polynucleotides of the invention can be incorporated into a recombinant replicable vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus

- 13 -

in a further embodiment, the invention provides a method of making polynucleotides of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from 5 the host cell. Suitable host cells are described below in connection with expression vectors.

Expression Vectors.

Preferably, a polynucleotide of the invention in a vector is operably linked to a regulatory sequence which is capable of providing for the expression of the coding sequence 10 by the host cell, i.e. the vector is an expression vector. The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A regulatory sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under condition compatible with the control sequences.

15 The term "regulatory sequences" includes promoters and enhancers and other expression regulation signals. These may be selected to be compatible with the host cell for which the expression vector is designed. For example, yeast-regulatory sequences include *S. cerevisiae* GAL4 and ADH promoters, *S. pombe* nmt1 and adh promoters. Mammalian promoters, such as  $\alpha$ -actin promoters, may be used. Mammalian promoters also include the 20 metallothionein promoter which can upregulate expression in response to heavy metals such as cadmium and is thus an inducible promoter. Tissue-specific promoters, for example neuronal cell specific may be used. Viral promoters may also be used, for example the Moloney murine leukaemia virus long terminal repeat (MMLV LTR), the promoter rous sarcoma virus (RSV) LTR promoter, the SV40 promoter, the human cytomegalovirus 25 (CMV) IE promoter, herpes simplex virus promoters or adenovirus promoters. All these promoters are readily available in the art.

Such vectors may be transformed into a suitable host cell as described above to provide for expression of a polypeptide of the invention. Thus, in a further aspect the invention provides a process for preparing polypeptides according to the invention which 30 comprises cultivating a host cell transformed or transfected with an expression vector as described above under conditions to provide for expression by the vector of a coding

- 14 -

sequence encoding the polypeptides, and recovering the expressed polypeptides.

The vectors may be for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of the said polynucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid or a neomycin resistance gene for a mammalian vector. Vectors may be used *in vitro*, for example for the production of RNA or used to transfet or transform a host cell. The vector may also be adapted to be used *in vivo*, for example in a method of gene therapy.

A further embodiment of the invention provides host cells transformed or transfected with the vectors for the replication and expression of polynucleotides of the invention. The cells will be chosen to be compatible with the said vector and may for example be bacterial, yeast, insect or mammalian.

Polynucleotides according to the invention may also be inserted into the vectors described above in an antisense orientation in order to provide for the production of antisense RNA. Antisense RNA or other antisense polynucleotides may also be produced by synthetic means. Such antisense polynucleotides may be used in a method of controlling the levels of MNGEF or its variants or species homologues.

#### D. Antibodies.

The invention also provides monoclonal or polyclonal antibodies to polypeptides of the invention or fragments thereof. The invention further provides a process for the production of monoclonal or polyclonal antibodies to polypeptides of the invention. Monoclonal antibodies may be prepared by conventional hybridoma technology using the polypeptides of the invention or peptide fragments thereof, as immunogens. Polyclonal antibodies may also be prepared by conventional means which comprise inoculating a host animal, for example a rat or a rabbit, with a polypeptide of the invention or peptide fragment thereof and recovering immune serum. In order that such antibodies may be made, the invention also provides polypeptides of the invention or fragments thereof haptenised to another polypeptide for use as immunogens in animals or humans.

For the purposes of this invention, the term "antibody", unless specified to the contrary, includes fragments of whole antibodies which retain their binding activity for a

- 15 -

tumour target antigen. Such fragments include Fv, F(ab') and F(ab')<sub>2</sub> fragments, as well as single chain antibodies. Furthermore, the antibodies and fragments thereof may be humanised antibodies, e.g. as described in EP-A-239400.

Antibodies may be used in method of detecting polypeptides of the invention  
5 present in biological samples by a method which comprises:

- (a) providing an antibody of the invention;
- (b) incubating a biological sample with said antibody under conditions which allow for the formation of an antibody-antigen complex; and
- (c) determining whether antibody-antigen complex comprising said antibody is formed.

Suitable samples include extracts from brain tissue, both normal and neoplastic. Suitable samples may also include extracts from other tissues such as breast, ovary, lung, colon, pancreas, testes, liver, muscle and bone tissues or from neoplastic growths derived from such tissues.

15 Antibodies of the invention may be bound to a solid support and/or packaged into kits in a suitable container along with suitable reagents, controls, instructions and the like.

#### E. Therapeutic uses

G-protein mediated signal transduction pathways have been shown to be involved in the control of cell division and growth. Many of the gene products involved in such pathways are proto-oncogenes i.e. they are capable of causing cellular transformation if mutated or aberrantly expressed, for example over-expressed. Therefore, mutations in MNGEF or its homologues may be a cause of cellular transformation, especially in the case of tumours associated with neuronal tissue, more particularly brain tissue. It may be possible to treat tumours that arise as a result by restoring normal MNGEF/NGEF function. This may be performed by means of gene therapy. Alternatively, it may be possible to raise antibodies that recognise specifically, mutated regions of the MNGEF protein, or its human homologue, NGEF. Such antibodies could be linked to therapeutic agents which would then target specifically cancer cells containing the mutated form of MNGEF/NGEF.

30 Thus the polypeptides, polynucleotides and antibodies of the invention may be used in as compounds for treating neoplasms in animals or humans. Typically the compounds

- 16 -

are formulated for clinical administration by mixing them with a pharmaceutically acceptable carrier or diluent. For example they can be formulated for topical, parenteral, intravenous, intramuscular, subcutaneous, intraocular or transdermal administration. Preferably, the compound is used in an injectable form. Direct injection into the patient's tumour is advantageous because it makes it possible to concentrate the therapeutic effect at the level of the affected tissues. It may therefore be mixed with any vehicle which is pharmaceutically acceptable for an injectable formulation, preferably for a direct injection at the site to be treated. The pharmaceutically carrier or diluent may be, for example, sterile or isotonic solutions.

The dose of compound used may be adjusted according to various parameters, especially according to the compound used, the age, weight and condition of the patient to be treated, the mode of administration used, pathology of the tumour and the required clinical regimen. As a guide, the amount of compound administered by injection is suitably from 0.01 mg/kg to 30 mg/kg, preferably from 0.1 mg/kg to 10 mg/kg.

The routes of administration and dosages described are intended only as a guide since a skilled practitioner will be able to determine readily the optimum route of administration and dosage for any particular patient and condition.

Compounds to be administered may include polypeptides, nucleic acids or antibodies. The nucleic acids may encode polypeptides or they may encode antisense constructs that inhibit expression of a cellular gene. Nucleic acids may be administered by, for example, lipofection or by viral vectors. For example, the nucleic acid may form part of a viral vector such as an adenovirus. When viral vectors are used, in general the dose administered is between  $10^4$  and  $10^{14}$  pfu/ml, preferably  $10^6$  to  $10^{10}$  pfu/ml. The term pfu ("plaque forming unit") corresponds to the infectivity of a virus solution and is determined by infecting an appropriate cell culture and measuring, generally after 48 hours, the number of plaques of infected cells. The techniques for determining the pfu titre of a viral solution are well documented in the literature.

Any cancer types may be treated by these methods, for example leukaemias, and solid tumours such as breast, ovary, lung, colon, pancreas, testes, liver, brain, muscle and bone tumour. Preferably, the tumour is a tumour of the nervous system, in particular the central nervous system, for example the brain.

- 17 -

The observation that MNGEF is expressed predominantly in brain tissue and that expression levels vary during foetal brain development (see Example 2) also suggest that MNGEF plays a role in neurological function, in particular neurological development. Thus it may be possible to diagnose, in particular prenatally, neurological conditions in which  
5 MNGEF and its human homologues are implicated using the detection methods discussed above. It may also be possible to treat such disorders by, in particular, gene therapy.

Mapping data indicate that MNGEF maps to mouse chromosome 1 within a region syntenic to human chromosome 2q. NGEF maps to human chromosome 2 by hybridisation to a panel of mono-chromosomal somatic cell hybrids. A form of the neurological disorder  
10 dystonia also maps to the long arm of human chromosome 2. Thus, human NGEF may be implicated in this disease. Therefore the above-mentioned probes and DNA sequences may be used to detect and diagnose dystonia in humans by, for example, determining the presence of mutant human NGEF sequences as described above. Alternatively, the gene encoding human NGEF may lie in close proximity to the gene implicated in a form of  
15 dystonia which maps to the long arm of human chromosome 2. Therefore the above-mentioned probes and DNA sequences may be used to detect and diagnose dystonia in humans by, for example, genetic linkage analysis using techniques well-known in the art including analysis of restriction fragment length polymorphisms associated with the human NGEF locus. Detection and diagnosis in both cases outlined above may be carried out pre-  
20 nately using foetal tissue, or extracts thereof, or post-natally. Detection and diagnosis may also be carried out on germline tissue or extracts thereof.

The following examples illustrate the invention:

25 **EXAMPLE 1 - Isolation of MNGEF2 and overlapping clones**

MNGEF2 and the overlapping clones were isolated from an adult mouse brain cDNA library (Izap Stratagene) cloned into the *EcoRI* and *XhoI* site of the vector pBluescript KS.

Approximately,  $10^6$  plaques were screened using a oligonucleotide designated  
30 M3/6T7 Forward from the M3/6 gene (5'GCAGGAAAGCTGGCAGCT 3' - SEQ ID No. 9). The probe was end-labelled with  $\gamma$ -<sup>32</sup>P dCTP (3000 Ci/mmol) using Promega

- 18 -

kinase. The MNGEF1, MNGEF2 and MNGEF3 cDNA clones were isolated from the host bacteriophage using a standard *in vivo* excision protocol. The three inserts were released from the vector by digestion with the restriction enzymes *EcoRI* and *XbaI*. The sizes of the MNGEF1, MNGEF2 and MNGEF3 clones were approximately 2.3, 2.8 and 1.35 kb  
5 respectively.

The clones were sequenced using a standard sequencing protocol from USB (Amersham). The full length cDNAs were digested using *TaqI* restriction enzyme and the resulting fragments were subcloned into the *ClaI* site of the vector pBluescript KS to facilitate sequencing. Full length sequencing in one direction was obtained by carrying out sequential walks using insert specific oligonucleotides. Sequence analysis was done using  
10 the GCG Wisconsin package version 8.

### Results

Approximately one million plaques from an adult mouse brain cDNA library were  
15 screened with an oligonucleotide (M3/6T7 Forward) from the M3/6 cDNA sequence. Five positives clones were identified, three of which appeared to be the same transcript of varying length. Sequencing of these cDNA clones demonstrated that they showed significant homology to TIM, a transforming gene, whose sequence is related to regulators  
20 of small GTP-binding proteins. 60% homology was observed on the nucleotide level between the MNGEF2 and TIM. The homology extended over the region known as DH domain, which plays an important role in mediating cellular transformation. Sequencing also revealed that two of these cDNA clones (MNGEF1 and MNGEF2) contained the following trinucleotide repeat (AGG)<sub>8</sub>GAG(AGG)<sub>3</sub> (SEQ ID No. 10). In addition it was observed that the longer of these cDNAs, MNGEF2, contained an extra 92bp sequence,  
25 which was not present in MNGEF1 and MNGEF3, although the flanking sequence of the region was identical. This 92 bp fragment comprises an unspliced intron which results in a premature termination codon as shown in SEQ ID NO. 3.

### **EXAMPLE 2 - Expression of MNGEF2 in mouse and human tissues**

30 To determine the pattern of expression of MNGEF, the cDNA clone MNGEF2 was hybridised to Northern blots of poly(A)+ RNA derived from a selection of adult mouse

- 19 -

tissues and human foetal brain tissues.

#### Northern analysis

RNA was extracted from mouse tissue and poly(A)+ RNA was prepared from  
5 100 µg of total RNA using the Dynabeads mRNA purification kit (Dynal). Northern blots  
were prepared according to Current Protocols in Molecular Biology, with each lane  
containing 2 µg of poly(A)+ RNA. The human foetal brain Northern blot and the mouse  
foetal developmental Northern blot were obtained from Clontech. The blots were  
hybridised at 42°C in standard formamide buffer and washed to a stringency of 0.1xSSC,  
10 0.1% SDS at 65°C. The blots were visualised by autoradiography after exposure for one or  
two days at -70°C.

#### Results

The MNGEF2 cDNA clone detected a transcript of approximately 3 kb  
15 predominantly in mouse brain and a faint one of the same size in mouse eye. In addition, a  
shorter transcript (approximately 2.2 kb) of less intensity was seen in the brain. A faint  
slightly larger transcript (about 3.5 kb) was also observed in small intestine and liver.

Hybridisation of the MNGEF2 cDNA clone to a Northern blot of human brain  
tissues (Clontech), detects a 3 kb transcript expressed predominantly in the caudate nucleus,  
20 but also in the amygdala and the hippocampus. The same sized transcript, albeit much  
fainter, was observed in all the remaining tissues.

A similar 3 kb transcript was seen when the MNGEF2 cDNA clone was used as a  
probe on a whole mouse embryo developmental Northern (Clontech). The strongest signal  
was observed in day 7 of embryonic development. Weaker signals of the same size were  
25 seen in days 11, 15 and 17.

#### **EXAMPLE 3 - Partial cloning of human NGEF**

To isolate the human homologue of MNGEF, primers m32bt7f and m32bt3f were  
used to amplify cDNA from human foetal brain. The sequences of the primers used are  
30 shown below:

3.2AT3F: 5'-CAAGAGAGGCTGGCAGAGGCAC-3' - SEQ ID No. 11

- 20 -

- 3.2AT7F: 5'-GGACCAAGTTGTATCCTTCAC-3' - SEQ ID No. 12  
3.2BT7F: 5'-GGACATCTGCTGCAGCTCAC-3' - SEQ ID No. 13  
3.2BT3F: 5'-GGAGAGCTCTGCCTCAGATCTG-3' - SEQ ID No. 14

5 An 803 bp product was amplified and cloned into the pGEMT vector (Promega).

The clone HFB32 was sequenced and the sequence is shown as SEQ ID No. 7. The translated protein sequence is shown as SEQ ID No. 8. A comparison between mouse and the human nucleotide sequence indicates 87.8% homology. A comparison between the protein sequence of the two species indicates 97% homology.

10 A search of the Yeast Genome database with the DH region of MNGEF showed homology to an open reading frame (ORF) from Chromosome XII (figure 6). This ORF corresponds to a yeast protein called ROM2 which is a GDP-GTP exchange protein for Rho1p containing the DH domains and the pleckstrin domains. The RHO1 gene encodes a homologue of the mammalian RhoA small GTP binding protein in yeast. Rho1p is localised at the growth site and required for bud formation. Disruption of ROM2 results in a temperature-sensitive growth phenotype. These mutants offer an attractive system to 15 study activation of Rho.

- 21 -

## SEQUENCE LISTING

(1ii) NUMBER OF SEQUENCES: 8

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2741 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 343..2004

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GCGCTCTACA GCAGCGGCGG CGGCAGCTCC GGCTTGAGCC GCGCGCGCTG CGACCTCACT	60
CAGAGCCCCGC GCATTGCCCG CGGCTGGGCC CTGGGCCCCG CGCGGCTCCC CACCAGCCCC	120
TGAGCCTTACCG CGGTGCTGG TCCCCATGGA GCTGCTGGCT GCAGCCTTCA GCGCCGCTG	180
CGCCGTGGAC CACGACAGCT CCACCTCGGA GAGCGACACG CGCGACTCGG CGGCGGGACA	240
CCTGCCGGGC AGCGAGTCAT CCTCCACCCCC TGAAATGGA ACCACACCCG AGGAGTGGCC	300
AGCCCTCACCG GACAGCCCCA CCACCTCTCAC GGAGCCCTGC AG ATG ATC CAT CCC Met Ile His Pro	354
1	
ATT CCC GCC GAC TCC TGG AGA AAC CTC ATT GAA CAA ATA GGG CTC CTG Ile Pro Ala Asp Ser Trp Arg Asn Leu Ile Glu Gln Ile Gly Leu Leu	402
5 10 15 20	
TAT CAA GAG TAT AGA GAC AAA TCG ACT CTC CAA GAA ATT GAA ACA CGG Tyr Gln Glu Tyr Arg Asp Lys Ser Thr Leu Gln Glu Ile Glu Thr Arg	450
25 30 35	
AGG CAG CAG GAT GCA GAA ATC CAA GGC AAC TCC GAT GGG TCC CAG GTT Arg Gln Gln Asp Ala Glu Ile Gln Gly Asn Ser Asp Gly Ser Gln Val	498
40 45 50	
GGG GAA GAC GCT GGA GAG GAG GAG GAG GAG GAG GGA GAG GAG Gly Glu Asp Ala Gly Glu Glu Glu Glu Glu Glu Glu Glu Glu	546
55 60 65	

-22-

GAG GAG CTG GCC AGC CCT CCT GAG AGG AGA GCT CTG CCT CAG ATC TGC 70 Glu Glu Leu Ala Ser Pro Pro Glu Arg Arg Ala Leu Pro Gln Ile Cys 75 80	594
CTG CTC AGC AAC CCC CAC TCC AGG TTC AAC CTC TGG CAA GAC CTT CCT 85 Leu Leu Ser Asn Pro His Ser Arg Phe Asn Leu Trp Gln Asp Leu Pro 90 95 100	642
GAG ATC CAG AGC AGT GGC GTG CTG GAC ATT CTC CAG CCG GAG GAG ATC Glu Ile Gln Ser Ser Gly Val Leu Asp Ile Leu Gln Pro Glu Glu Ile 105 110 115	690
AGG CTG CAG GAG GCC ATG TTT GAG TTG GTT ACC TCT GAG GCC TCC TAC Arg Leu Gln Glu Ala Met Phe Glu Leu Val Thr Ser Glu Ala Ser Tyr 120 125 130	738
TAT AAG AGC CTG AAC CTG CTG GTG TCG CAC TTC ATG GAG AAC GAG CGT Tyr Lys Ser Leu Asn Leu Leu Val Ser His Phe Met Glu Asn Glu Arg 135 140 145	786
CTG AAG AAG ATC CTG CAT CCA TCT GAG GCC CAC ATC CTC TTT TCC AAT Leu Lys Lys Ile Leu His Pro Ser Glu Ala His Ile Leu Phe Ser Asn 150 155 160	834
GTC CTG GAT GTC ATG GCT GTC AGT GAG CGG TTT TTG CTG GAG CTA GAG Val Leu Asp Val Met Ala Val Ser Glu Arg Phe Leu Leu Glu Leu Glu 165 170 175 180	882
CAC CGC ATG GAG GAG AAC ATT GTT ATC TCG GAT GTG TGC GAC ATC GTG His Arg Met Glu Glu Asn Ile Val Ile Ser Asp Val Cys Asp Ile Val 185 190 195	930
TAC CGT TAC GCA GCT GAT CAC TTC TCG GTC TAT ATC ACT TAC GTC AGT Tyr Arg Tyr Ala Ala Asp His Phe Ser Val Tyr Ile Thr Tyr Val Ser 200 205 210	978
AAC CAG ACC TAC CAG GAA AGG ACA TAC AAG CAG CTC CTA CAG GAG AAG Asn Gln Thr Tyr Gln Glu Arg Thr Tyr Lys Gln Leu Leu Gln Glu Lys 215 220 225	1026
GCC GCT TTC CCG GAA CTG ATC GCA CAG TTG GAG CTG GAC CCC AAA TGC Ala Ala Phe Arg Glu Leu Ile Ala Gln Leu Glu Leu Asp Pro Lys Cys 230 235 240	1074
AAG GGC CTG CCT TTC TCC TCC TTC CTC ATC TTG CCT TTC CAG AGG ATC Lys Gly Leu Pro Phe Ser Ser Phe Leu Ile Leu Pro Phe Gln Arg Ile 245 250 255 260	1122

- 23 -

ACG AGA CTC AAG CTG CTG GTC CAG AAT ATC CTG AAG AGA GTG GAG GAG Thr Arg Leu Lys Leu Leu Val Glu Asn Ile Leu Lys Arg Val Glu Glu 265 270 275	1170
AGG TCT GAA CGT GAA GGC ACC GCC TTG GAT GCC CAC AAG GAG CTA GAA Arg Ser Glu Arg Glu Gly Thr Ala Leu Asp Ala His Lys Glu Leu Glu 280 285 290	1218
ATG GTG GTA AAG GCA TGC AAT GAG GGT GTC CGG AAG ATG AGC CGC ACA Met Val Val Lys Ala Cys Asn Glu Gly Val Arg Lys Met Ser Arg Thr 295 300 305	1266
GAA CAG ATG ATC AGC ATT CAG AAG AAG ATG GAG TTC AAG ATC AAG TCG Glu Gln Met Ile Ser Ile Gln Lys Lys Met Glu Phe Lys Ile Lys Ser 310 315 320	1314
GTA CCC ATC ATC TCA CAC TCC CGG TGG CTG CTG AAG CAG GGT GAG CTG Val Pro Ile Ile Ser His Ser Arg Trp Leu Leu Lys Gln Gly Glu Leu 325 330 335 340	1362
CAG CAG ATG TCC GGC CCC AAG ACC TCC CGC ACC CTG CGG ACC AAG AAG Gln Gln Met Ser Gly Pro Lys Thr Ser Arg Thr Leu Arg Thr Lys Lys 345 350 355	1410
CTC TTC AGA GAA ATT TAC CTC TTC CTC TTC AAT GAC CTG CTG GTG ATC Leu Phe Arg Glu Ile Tyr Leu Phe Leu Phe Asn Asp Leu Leu Val Ile 360 365 370	1458
TGC CGG CAG ATC CCT GGA GAC AAG TAC CAG GTG TTT GAT TCG GCC CCA Cys Arg Gln Ile Pro Gly Asp Lys Tyr Gln Val Phe Asp Ser Ala Pro 375 380 385	1506
AGG GGC CTG CTT CGA GTG GAG GAG CTG GAG GAC CAG GGT CAA ACA CTG Arg Gly Leu Leu Arg Val Glu Glu Leu Glu Asp Gln Gly Gln Thr Leu 390 395 400	1554
GCT AAT GTG TTC ATC CTG CGG CTG CTG GAA AAT GCA GAT GAC CGA GAG Ala Asn Val Phe Ile Leu Arg Leu Leu Glu Asn Ala Asp Asp Arg Glu 405 410 415 420	1602
GCC ACC TAT ATG CTG AAG GCA TCC TCC CAG AGC GAG ATG AAG CGC TGG Ala Thr Tyr Met Leu Lys Ala Ser Ser Gln Ser Glu Met Lys Arg Trp 425 430 435	1650
ATG ACC TCA CTG GCC CCC AAC AGG AGG ACC AAG TTT GTA TCC TTC ACA Met Thr Ser Leu Ala Pro Asn Arg Arg Thr Lys Phe Val Ser Phe Thr 440 445 450	1698

- 24 -

TCT CGG CTG TTG GAC TGT CCC CAG GTC CAG TGT GTG CAC CCG TAT GTG Ser Arg Leu Leu Asp Cys Pro Gln Val Gln Cys Val His Pro Tyr Val	1746
455 460 465	
GCC CAG CAG CCT GAT GAA CTG ACG CTG GAA CTG GCA GAT ATC CTG AAC Ala Gln Gln Pro Asp Glu Leu Thr Leu Glu Leu Ala Asp Ile Leu Asn	1794
470 475 480	
ATC CTG GAG AAG ACA GAG GAT GGG TGG ATC TTT GGT GAG CGG CTG CAT Ile Leu Glu Lys Thr Glu Asp Gly Trp Ile Phe Gly Glu Arg Leu His	1842
485 490 495 500	
GAC CAG GAG AGA GGC TGG TTC CCC AGT TCC ATG ACA GAG GAG ATC CTG Asp Gln Glu Arg Gly Trp Phe Pro Ser Ser Met Thr Glu Glu Ile Leu	1890
505 510 515	
AAC CCC AAG ATC CGC TCC CAG AAC CTC AAG GAA TGT TTC CGG GTA CAT Asn Pro Lys Ile Arg Ser Gln Asn Leu Lys Glu Cys Phe Arg Val His	1938
520 525 530	
AAG ATG GAA GAC CCT CAG CGC AGC CAG AAT AAG GAC CGC AGG AAG CTG Lys Met Glu Asp Pro Gln Arg Ser Gln Asn Lys Asp Arg Arg Lys Leu	1986
535 540 545	
GGC AGC CGG AAT CGT CAA TGAAACCTCCC CAGCTCAGGC ACCTGAAGGG Gly Ser Arg Asn Arg Gln	2034
550	
AAGGGTGTGG GCAGGGATGG GGAGCAGGCC CGGCAGAGAC GCCCGACAGA TTCAAGAGGC	2094
CTTAGGGAAG AATGTCAGTG CCTTCTCAGG CAGCAGGAGT GGCTTCGGCC TGCTCTGTCC	2154
CTGCCCATGC TGTGGAAGCT CTAGTGTCTT GGCCACTTGT TTGCTTGAC ACTGGTAAA	2214
AGCTAAGTAC TTAGGCAGTA TTACACCACC TCCCTTCAGT CTCTCAGAGG TAGAAGAAGG	2274
CAGGCATGCT CCAGAGACCT TCCGGTGACT GGAAGAGGCC CACACAAGGG TCCCTGGCAG	2334
CAGGCAGGTG GAAGGTAACC ACTGTCAGGA TCCCCTGAAC TGCACTGTGTC CTTCCCTACT	2394
TTGGAAGCTG TTAAGAGTCT ACCAGGCACA CAGATGGCCG CCCCTGCCG AGGGAGTTG	2454
ATGAGCAGTG GTGACCCCTGC CTGCCGTCC CCGTGCCTCT GCCAGCCTCT CTTGCACGCC	2514
AAGCCCTGCC CTCAGCAGGC TTCCCAAAGC TTAGCTGAGG GTTCATGCCA CCTCTAGCTC	2574
CTTGAAGGGC TTGATATCAC TTGTGTCTCC TGGGCCCCCTG ATGGAGCCCA GGCGTTTGC	2634
AGAATGAATT GGTCACTGCA TCCTTATGG TCATGGTTT GAGAAAAGCA AATATCATT	2694

- 25 -

TTGGCTGCAT TAAAAGAAGC ATCCTATATA AAAAAAAA AAAAAAA

2741

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 554 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Ile His Pro Ile Pro Ala Asp Ser Trp Arg Asn Leu Ile Glu Gln  
1 5 10 15

Ile Gly Leu Leu Tyr Gln Glu Tyr Arg Asp Lys Ser Thr Leu Gln Glu  
20 25 30

Ile Glu Thr Arg Arg Gln Gln Asp Ala Glu Ile Gln Gly Asn Ser Asp  
35 40 45

Gly Ser Gln Val Gly Glu Asp Ala Gly Glu Glu Glu Glu Glu Glu  
50 55 60

Glu Gly Glu Glu Glu Glu Leu Ala Ser Pro Pro Glu Arg Arg Ala Leu  
65 70 75 80

Pro Gln Ile Cys Leu Leu Ser Asn Pro His Ser Arg Phe Asn Leu Trp  
85 90 95

Gln Asp Leu Pro Glu Ile Gln Ser Ser Gly Val Leu Asp Ile Leu Gln  
100 105 110

Pro Glu Glu Ile Arg Leu Gln Glu Ala Met Phe Glu Leu Val Thr Ser  
115 120 125

Glu Ala Ser Tyr Tyr Lys Ser Leu Asn Leu Leu Val Ser His Phe Met  
130 135 140

Glu Asn Glu Arg Leu Lys Lys Ile Leu His Pro Ser Glu Ala His Ile  
145 150 155 160

Leu Phe Ser Asn Val Leu Asp Val Met Ala Val Ser Glu Arg Phe Leu  
165 170 175

Leu Glu Leu Glu His Arg Met Glu Glu Asn Ile Val Ile Ser Asp Val  
180 185 190

-26-

Cys Asp Ile Val Tyr Arg Tyr Ala Ala Asp His Phe Ser Val Tyr Ile  
195 200 205

Thr Tyr Val Ser Asn Gln Thr Tyr Gln Glu Arg Thr Tyr Lys Gln Leu  
210 215 220

Leu Gln Glu Lys Ala Ala Phe Arg Glu Leu Ile Ala Gln Leu Glu Leu  
225 230 235 240

Asp Pro Lys Cys Lys Gly Leu Pro Phe Ser Ser Phe Leu Ile Leu Pro  
245 250 255

Phe Gln Arg Ile Thr Arg Leu Lys Leu Leu Val Gln Asn Ile Leu Lys  
260 265 270

Arg Val Glu Glu Arg Ser Glu Arg Glu Gly Thr Ala Leu Asp Ala His  
275 280 285

Lys Glu Leu Glu Met Val Val Lys Ala Cys Asn Glu Gly Val Arg Lys  
290 295 300

Met Ser Arg Thr Glu Gln Met Ile Ser Ile Gln Lys Lys Met Glu Phe  
305 310 315 320

Lys Ile Lys Ser Val Pro Ile Ile Ser His Ser Arg Trp Leu Leu Lys  
325 330 335

Gln Gly Glu Leu Gln Gln Met Ser Gly Pro Lys Thr Ser Arg Thr Leu  
340 345 350

Arg Thr Lys Lys Leu Phe Arg Glu Ile Tyr Leu Phe Leu Phe Asn Asp  
355 360 365

Leu Leu Val Ile Cys Arg Gln Ile Pro Gly Asp Lys Tyr Gln Val Phe  
370 375 380

Asp Ser Ala Pro Arg Gly Leu Leu Arg Val Glu Glu Leu Glu Asp Gln  
385 390 395 400

Gly Gln Thr Leu Ala Asn Val Phe Ile Leu Arg Leu Leu Glu Asn Ala  
405 410 415

Asp Asp Arg Glu Ala Thr Tyr Met Leu Lys Ala Ser Ser Gln Ser Glu  
420 425 430

Met Lys Arg Trp Met Thr Ser Leu Ala Pro Asn Arg Arg Thr Lys Phe  
435 440 445

Val Ser Phe Thr Ser Arg Leu Leu Asp Cys Pro Gln Val Gln Cys Val  
450 455 460

- 27 -

His Pro Tyr Val Ala Gln Gln Pro Asp Glu Leu Thr Leu Glu Leu Ala  
 465                    470                    475                    480  
 Asp Ile Leu Asn Ile Leu Glu Lys Thr Glu Asp Gly Trp Ile Phe Gly  
 485                    490                    495  
 Glu Arg Leu His Asp Gln Glu Arg Gly Trp Phe Pro Ser Ser Met Thr  
 500                    505                    510  
 Glu Glu Ile Leu Asn Pro Lys Ile Arg Ser Gln Asn Leu Lys Glu Cys  
 515                    520                    525  
 Phe Arg Val His Lys Met Glu Asp Pro Gln Arg Ser Gln Asn Lys Asp  
 530                    535                    540  
 Arg Arg Lys Leu Gly Ser Arg Asn Arg Gln  
 545                    550

## (2) INFORMATION FOR SEQ ID NO: 3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2833 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 343..1860

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGCTCTACA GCAGCGCGG CGGCAGCTCC GGCTTGAGCC GCGCGCGCTG CGACCTCACT	60
CAGAGCCCGC GCATTGCCCG CGGCTGGGCC CTGGGCCCCG CGCGGCTCCC CACCAGCCCC	120
TGAGCCTACC CGGTCGCTGG TCCCCATGGA GCTGCTGGCT GCAGCCTCA GCGCCGCGCTG	180
CGCCGTGGAC CACGACAGCT CCACCTCGGA GAGCGACACG CGCGACTCGG CGGCGGGACA	240
CCTGCCGGGC AGCGAGTCAT CCTCCACCCCG TGAAATGGA ACCACACCCG AGGAGTGC	300
AGCCCTCACC GACAGCCCCA CCACCTCAC GGAGCCCTGC AG ATG ATC CAT CCC Met Ile His Pro	354

- 28 -

ATT CCC GCC GAC TCC TGG AGA AAC CTC ATT GAA CAA ATA GGG CTC CTG Ile Pro Ala Asp Ser Trp Arg Asn Leu Ile Glu Gln Ile Gly Leu Leu	5	10	15	20	402
TAT CAA GAG TAT AGA GAC AAA TCG ACT CTC CAA GAA ATT GAA ACA CGG Tyr Gln Glu Tyr Arg Asp Lys Ser Thr Leu Gln Glu Ile Glu Thr Arg	25	30	35		450
AGG CAG CAG GAT GCA GAA ATC CAA GGC AAC TCC GAT GGG TCC CAG GTT Arg Gln Gln Asp Ala Glu Ile Gln Gly Asn Ser Asp Gly Ser Gln Val	40	45	50		498
GGG GAA GAC GCT GGA GAG GAG GAG GAG GAG GAG GGA GAG GAG Gly Glu Asp Ala Gly Glu Glu Glu Glu Glu Glu Glu Glu Gly Glu Glu	55	60	65		546
GAG GAG CTG GCC AGC CCT CCT GAG AGG AGA GCT CTG CCT CAG ATC TGC Glu Glu Leu Ala Ser Pro Pro Glu Arg Arg Ala Leu Pro Gln Ile Cys	70	75	80		594
CTG CTC AGC AAC CCC CAC TCC AGG TTC AAC CTC TGG CAA GAC CTT CCT Leu Leu Ser Asn Pro His Ser Arg Phe Asn Leu Trp Gln Asp Leu Pro	85	90	95	100	642
GAG ATC CAG AGC AGT GGC GTG CTG GAC ATT CTC CAG CCG GAG GAG ATC Glu Ile Gln Ser Ser Gly Val Leu Asp Ile Leu Gln Pro Glu Glu Ile	105	110	115		690
AGG CTG CAG GAG GCC ATG TTT GAG TTG GTT ACC TCT GAG GCC TCC TAC Arg Leu Gln Ala Met Phe Glu Leu Val Thr Ser Glu Ala Ser Tyr	120	125	130		738
TAT AAG AGC CTG AAC CTG CTG GTG TCG CAC TTC ATG GAG AAC GAG CGT Tyr Lys Ser Leu Asn Leu Leu Val Ser His Phe Met Glu Asn Glu Arg	135	140	145		786
CTG AAG AAG ATC CTG CAT CCA TCT GAG GCC CAC ATC CTC TTT TCC AAT Leu Lys Lys Ile Leu His Pro Ser Glu Ala His Ile Leu Phe Ser Asn	150	155	160		834
GTC CTG GAT GTC ATG GCT GTC AGT GAG CGG TTT TTG CTG GAG CTA GAG Val Leu Asp Val Met Ala Val Ser Glu Arg Phe Leu Leu Glu Leu Glu	165	170	175	180	882
CAC CGC ATG GAG GAG AAC ATT GTT ATC TCG GAT GTG TGC GAC ATC GTG His Arg Met Glu Glu Asn Ile Val Ile Ser Asp Val Cys Asp Ile Val	185	190	195		930

- 29 -

TAC CGT TAC GCA GCT GAT CAC TTC TCG GTC TAT ATC ACT TAC GTC AGT Tyr Arg Tyr Ala Ala Asp His Phe Ser Val Tyr Ile Thr Tyr Val Ser 200 205 210	978
AAC CAG ACC TAC CAG GAA AGG ACA TAC AAG CAG CTC CTA CAG GAG AAG Asn Gln Thr Tyr Gln Glu Arg Thr Tyr Lys Gln Leu Leu Gln Glu Lys 215 220 225	1026
GCC GCT TTC CGG GAA CTG ATC GCA CAG TTG GAG CTG GAC CCC AAA TGC Ala Ala Phe Arg Glu Leu Ile Ala Gln Leu Glu Leu Asp Pro Lys Cys 230 235 240	1074
AAG GGC CTG CCT TTC TCC TCC CTC ATC TTG CCT TTC CAG AGG ATC Lys Gly Leu Pro Phe Ser Ser Phe Leu Ile Leu Pro Phe Gln Arg Ile 245 250 255 260	1122
ACG AGA CTC AAG CTG CTG GTC CAG AAT ATC CTG AAG AGA GTG GAG GAG Thr Arg Leu Lys Leu Leu Val Gln Asn Ile Leu Lys Arg Val Glu Glu 265 270 275	1170
AGG TCT GAA CGT GAA GGC ACC GCC TTG GAT GCC CAC AAG GAG CTA GAA Arg Ser Glu Arg Glu Gly Thr Ala Leu Asp Ala His Lys Glu Leu Glu 280 285 290	1218
ATG GTG GTA AAG GCA TGC AAT GAG GGT GTC CGG AAG ATG AGC CGC ACA Met Val Val Lys Ala Cys Asn Glu Gly Val Arg Lys Met Ser Arg Thr 295 300 305	1266
GAA CAG ATG ATC AGC ATT CAG AAG AAG ATG GAG TTC AAG ATC AAG TCG Glu Gln Met Ile Ser Ile Gln Lys Lys Met Glu Phe Lys Ile Lys Ser 310 315 320	1314
GTA CCC ATC ATC TCA CAC TCC CGG TGG CTG CTG AAG CAG GGT GAG CTG Val Pro Ile Ile Ser His Ser Arg Trp Leu Leu Lys Gln Gly Glu Leu 325 330 335 340	1362
CAG CAG ATG TCC GGC CCC AAG ACC TCC CGC ACC CTG CGG ACC AAG AAG Gln Gln Met Ser Glu Pro Lys Thr Ser Arg Thr Leu Arg Thr Lys Lys 345 350 355	1410
CTC TTC AGA GAA ATT TAC CTC TTC CTC AAT GAC CTG CTG GTG ATC Leu Phe Arg Glu Ile Tyr Leu Phe Leu Phe Asn Asp Leu Leu Val Ile 360 365 370	1458
TGC CGG CAG ATC CCT GGA GAC AAG TAC CAG GTG TTT GAT TCG GCC CCA Cys Arg Gln Ile Pro Glu Asp Lys Tyr Gln Val Phe Asp Ser Ala Pro 375 380 385	1506

- 30 -

AGG GGC CTG CTT CGA GTG GAG GAG CTG GAG GAC CAG GGT CAA ACA CTG Arg Gly Leu Leu Arg Val Glu Glu Leu Glu Asp Gln Gly Gln Thr Leu 390 395 400	1554
GCT AAT GTG TTC ATC CTG CGG CTG CTG GAA AAT GCA GAT GAC CGA GAG Ala Asn Val Phe Ile Leu Arg Leu Leu Glu Asn Ala Asp Asp Arg Glu 405 410 415 420	1602
GCC ACC TAT ATG CTG AAG GCA TCC TCC CAG AGC GAG ATG AAG CGC TGG Ala Thr Tyr Met Leu Lys Ala Ser Ser Gln Ser Glu Met Lys Arg Trp 425 430 435	1650
ATG ACC TCA CTG GCC CCC AAC AGG AGG ACC AAG TTT GTA TCC TTC ACA Met Thr Ser Leu Ala Pro Asn Arg Arg Thr Lys Phe Val Ser Phe Thr 440 445 450	1698
TCT CGG CTG TTG GAC TGT CCC CAG GTC CAG TGT GTG CAC CCG TAT GTG Ser Arg Leu Leu Asp Cys Pro Gln Val Gln Cys Val His Pro Tyr Val 455 460 465	1746
GCC CAG CAG CCT GAT GAA CTG ACG CTG GAA CTG GCA GAT ATC CTG AAC Ala Gln Gln Pro Asp Glu Leu Thr Leu Glu Leu Ala Asp Ile Leu Asn 470 475 480	1794
ATC CTG GAG AAG ACA GAG GAT GGT GAG CCC CGC ACC AAG GGG ACT CTG Ile Leu Glu Lys Thr Glu Asp Gly Glu Pro Arg Thr Lys Gly Thr Leu 485 490 495 500	1842
CAT CTT GGC CAG CCA TGA GAGAGAGGAC TATGGCCTAG ATGTAGGACT His Leu Gly Gln Pro * 505	1890
AGATGGTGCA GTTAGCAGGG TGGATCTTG GTGAGCGGCT GCATGACCAAG GAGAGAGGCT	1950
GGTCCCGAG TTCCATGACA GAGGAGATCC TGAACCCCAA GATCCGCTCC CAGAACCTCA	2010
AGGAATGTTT CCGGGTACAT AAGATGGAAG ACCCTCAGCG CAGCCAGAAT AAGGACCGCA	2070
GGAAAGCTGGG CAGCCGGAAT CGTCAATGAA CCTCCCCAGC TCAGGCACCT GAAGGGAAAGG	2130
GTGTGGGCAG GGATGGGGAG CAGGCCGGC AGAGACGCC GACAGATTCA GAGGGCCTTA	2190
GGGAAGAATG TCAGTGCCTT CTCAGGCAGC AGGAGTGGCT TCGGCCTGCT CTGTCCCTGC	2250
CCATGCTGTG GAAGCTCTAG TGTCTGGCC ACTTGTTCGC TTGCACACTG GTGAAAAGCT	2310
AAGTACTTAG GCAGTATTAC ACCACCTCCC TTCAGTCTCT CAGAGGTAGA AGAAGGCAGG	2370
CATGCTCCAG AGACCTTCCG GTGACTGGAA GAGGCCACA CAAGGGTCCC TGGCAGCAGG	2430

- 31 -

CAGGTGGAAG GTAACCCTG TCAGGATCCC CTGAAC TGCA CGTGTCCCTTC CCTACTTTGG	2490
AAGCTGTTAA GAGTCTACCA GGCACACAGA TGGCCGCCCC TGCCCGAGGG AGTTTGATGA	2550
GCAGTGGTGA CCCTGCCTGC CCGTCCCCGT GCCTCTGCCA GCCTCTCTTG CACGCCAAGC	2610
CCTGCCCTCA GCAGGCTTCC CAAAGCTTAG CTGAGGCTTC ATCCCACCTG TAGCTCTTGC	2670
AAGGGCTTGA TATCACTTGT GTCTCCTGGG CCCCTGATGG AGCCCAGGCG TTTTGAGAA	2730
TGAATTGGTC ACTGCATCCT TTATGGTCAT GGTTTGAGA AAAGCAAATA TCATTTTG	2790
CTGCATTAAGAAGCATCC TATATAAAAAA AAAAAAAA AAA	2833

## (2) INFORMATION FOR SEQ ID NO: 4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 505 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Ile His Pro Ile Pro Ala Asp Ser Trp Arg Asn Leu Ile Glu Gln			
1	5	10	15
Ile Gly Leu Leu Tyr Gln Glu Tyr Arg Asp Lys Ser Thr Leu Gln Glu			
20	25	30	
Ile Glu Thr Arg Arg Gln Gln Asp Ala Glu Ile Gln Gly Asn Ser Asp			
35	40	45	
Gly Ser Gln Val Gly Glu Asp Ala Gly Glu Glu Glu Glu Glu Glu Glu			
50	55	60	
Glu Glu Glu Glu Glu Leu Ala Ser Pro Pro Glu Arg Arg Ala Leu			
65	70	75	80
Pro Gln Ile Cys Leu Leu Ser Asn Pro His Ser Arg Phe Asn Leu Trp			
85	90	95	
Gln Asp Leu Pro Glu Ile Gln Ser Ser Gly Val Leu Asp Ile Leu Gln			
100	105	110	
Pro Glu Glu Ile Arg Leu Gln Glu Ala Met Phe Glu Leu Val Thr Ser			
115	120	125	

-32-

Glu Ala Ser Tyr Tyr Lys Ser Leu Asn Leu Leu Val Ser His Phe Met  
 130 135 140

Glu Asn Glu Arg Leu Lys Lys Ile Leu His Pro Ser Glu Ala His Ile  
 145 150 155 160

Leu Phe Ser Asn Val Leu Asp Val Met Ala Val Ser Glu Arg Phe Leu  
 165 170 175

Leu Glu Leu Glu His Arg Met Glu Glu Asn Ile Val Ile Ser Asp Val  
 180 185 190

Cys Asp Ile Val Tyr Arg Tyr Ala Ala Asp His Phe Ser Val Tyr Ile  
 195 200 205

Thr Tyr Val Ser Asn Gln Thr Tyr Gln Glu Arg Thr Tyr Lys Gln Leu  
 210 215 220

Leu Gln Glu Lys Ala Ala Phe Arg Glu Leu Ile Ala Gln Leu Glu Leu  
 225 230 235 240

Asp Pro Lys Cys Lys Gly Leu Pro Phe Ser Ser Phe Leu Ile Leu Pro  
 245 250 255

Phe Gln Arg Ile Thr Arg Leu Lys Leu Leu Val Gln Asn Ile Leu Lys  
 260 265 270

Arg Val Glu Glu Arg Ser Glu Arg Glu Gly Thr Ala Leu Asp Ala His  
 275 280 285

Lys Glu Leu Glu Met Val Val Lys Ala Cys Asn Glu Gly Val Arg Lys  
 290 295 300

Met Ser Arg Thr Glu Gln Met Ile Ser Ile Gln Lys Lys Met Glu Phe  
 305 310 315 320

Lys Ile Lys Ser Val Pro Ile Ile Ser His Ser Arg Trp Leu Leu Lys  
 325 330 335

Gln Gly Glu Leu Gln Gln Met Ser Gly Pro Lys Thr Ser Arg Thr Leu  
 340 345 350

Arg Thr Lys Lys Leu Phe Arg Glu Ile Tyr Leu Phe Leu Asn Asp  
 355 360 365

Leu Leu Val Ile Cys Arg Gln Ile Pro Gly Asp Lys Tyr Gln Val Phe  
 370 375 380

Asp Ser Ala Pro Arg Gly Leu Leu Arg Val Glu Glu Leu Glu Asp Gln  
 385 390 395 400

- 33 -

Gly Gln Thr Leu Ala Asn Val Phe Ile Leu Arg Leu Leu Glu Asn Ala  
 405                          410                          415

Asp Asp Arg Glu Ala Thr Tyr Met Leu Lys Ala Ser Ser Gln Ser Glu  
 420                          425                          430

Met Lys Arg Trp Met Thr Ser Leu Ala Pro Asn Arg Arg Thr Lys Phe  
 435                          440                          445

Val Ser Phe Thr Ser Arg Leu Leu Asp Cys Pro Gln Val Gln Cys Val  
 450                          455                          460

His Pro Tyr Val Ala Gln Gln Pro Asp Glu Leu Thr Leu Glu Leu Ala  
 465                          470                          475                          480

Asp Ile Leu Asn Ile Leu Glu Lys Thr Glu Asp Gly Glu Pro Arg Thr  
 485                          490                          495

Lys Gly Thr Leu His Leu Gly Gln Pro \*  
 500                          505

## (2) INFORMATION FOR SEQ ID NO: 5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2343 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..1609

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

C CTG TAT CAA GAG TAT AGA GAC AAA TCG ACT CTC CAA GAA ATT GAA	46
Leu Tyr Gln Glu Tyr Arg Asp Lys Ser Thr Leu Gln Glu Ile Glu	
1                        5                        10                        15	
ACA CGG AGG CAG CAG GAT GCA GAA ATC CAA GGC AAC TCC GAT GGG TCC	94
Thr Arg Arg Gln Gln Asp Ala Glu Ile Gln Gly Asn Ser Asp Gly Ser	
20                        25                        30	

- 34 -

CAG GTT GGG GAA GAC GCT GGA GAG GAG GAG GAG GAG GAG GAG GGA Gln Val Gly Glu Asp Ala Gly Glu Glu Glu Glu Glu Glu Glu Gly	35	40	45	142
GAG GAG GAG GAG CTG GCC AGC CCT CCT GAG AGG AGA GCT CTG CCT CAG Glu Glu Glu Leu Ala Ser Pro Pro Glu Arg Arg Ala Leu Pro Gln	50	55	60	190
ATC TGC CTG CTC AGC AAC CCC CAC TCC AGG TTC AAC CTC TGG CAA GAC Ile Cys Leu Leu Ser Asn Pro His Ser Arg Phe Asn Leu Trp Gln Asp	65	70	75	238
CTT CCT GAG ATC CAG AGC AGT GGC GTG CTG GAC ATT CTC CAG CCG GAG Leu Pro Glu Ile Gln Ser Ser Gly Val Leu Asp Ile Leu Gln Pro Glu	80	85	90	286
GAG ATC AGG CTG CAG GAG GCC ATG TTT GAG TTG GTT ACC TCT GAG GCC Glu Ile Arg Leu Gln Glu Ala Met Phe Glu Leu Val Thr Ser Glu Ala	100	105	110	334
TCC TAC TAT AAG AGC CTG AAC CTG CTG GTG TCG CAC TTC ATG GAG AAC Ser Tyr Tyr Lys Ser Leu Asn Leu Leu Val Ser His Phe Met Glu Asn	115	120	125	382
GAG CGT CTG AAG AAG ATC CTG CAT CCA TCT GAG GCC CAC ATC CTC TTT Glu Arg Leu Lys Lys Ile Leu His Pro Ser Glu Ala His Ile Leu Phe	130	135	140	430
TCC AAT GTC CTG GAT GTC ATG GCT GTC AGT GAG CGG TTT TTG CTG GAG Ser Asn Val Leu Asp Val Met Ala Val Ser Glu Arg Phe Leu Leu Glu	145	150	155	478
CTA GAG CAC CGC ATG GAG AAC ATT GTT ATC TCG GAT GTG TGC GAC Leu Glu His Arg Met Glu Glu Asn Ile Val Ile Ser Asp Val Cys Asp	160	165	170	526
ATC GTG TAC CGT TAC GCA GCT GAT CAC TTC TCG GTC TAT ATC ACT TAC Ile Val Tyr Arg Tyr Ala Ala Asp His Phe Ser Val Tyr Ile Thr Tyr	180	185	190	574
GTC AGT AAC CAG ACC TAC CAG GAA AGG ACA TAC AAG CAG CTC CTA CAG Val Ser Asn Gln Thr Tyr Gln Glu Arg Thr Tyr Lys Gln Leu Leu Gln	195	200	205	622
GAG AAG GCC GCT TTC CCG GAA CTG ATC GCA CAG TTG GAG CTG GAC CCC Glu Lys Ala Ala Phe Arg Glu Leu Ile Ala Gln Leu Glu Leu Asp Pro	210	215	220	670

- 35 -

AAA TGC AAG GGC CTG CCT TTC TCC TCC TTC CTC ATC TTG CCT TTC CAG Lys Cys Lys Gly Leu Pro Phe Ser Ser Phe Leu Ile Leu Pro Phe Gln 225                230                235	718
AGG ATC ACG AGA CTC AAG CTG CTG GTC CAG AAT ATC CTG AAG AGA GTG Arg Ile Thr Arg Leu Lys Leu Leu Val Gln Asn Ile Leu Lys Arg Val 248                245                250                255	766
GAG GAG AGG TCT GAA CGT GAA GGC ACC GCC TTG GAT GCC CAC AAG GAG Glu Glu Arg Ser Glu Arg Glu Gly Thr Ala Leu Asp Ala His Lys Glu 260                265                270	814
CTA GAA ATG GTG GTA AAG GCA TGC AAT GAG GGT GTC CGG AAG ATG AGC Leu Glu Met Val Val Lys Ala Cys Asn Glu Gly Val Arg Lys Met Ser 275                280                285	862
CGC ACA GAA CAG ATG ATC AGC ATT CAG AAG AAG ATG GAG TTC AAG ATC Arg Thr Glu Gln Met Ile Ser Ile Gln Lys Lys Met Glu Phe Lys Ile 290                295                300	910
AAG TCG GTA CCC ATC ATC TCA CAC TCC CGG TGG CTG CTG AAG CAG GGT Lys Ser Val Pro Ile Ile Ser His Ser Arg Trp Leu Leu Lys Gln Gly 305                310                315	958
GAG CTG CAG CAG ATG TCC GGC CCC AAG ACC TCC CGC ACC CTG CGG ACC Glu Leu Gln Gln Met Ser Gly Pro Lys Thr Ser Arg Thr Leu Arg Thr 320                325                330                335	1006
AAG AAG CTC TTC AGA GAA ATT TAC CTC TTC CTC TTC AAT GAC CTG CTG Lys Lys Leu Phe Arg Glu Ile Tyr Leu Phe Leu Asn Asp Leu Leu 340                345                350	1054
GTG ATC TGC CGG CAG ATC CCT GGA GAC AAG TAC CAG GTG TTT GAT TCG Val Ile Cys Arg Gln Ile Pro Gly Asp Lys Tyr Gln Val Phe Asp Ser 355                360                365	1102
GCC CCA AGG GGC CTG CTT CGA GTG GAG GAG CTG GAG GAC CAG GGT CAA Ala Pro Arg Gly Leu Leu Arg Val Glu Glu Leu Glu Asp Gln Gly Gln 370                375                380	1150
ACA CTG GCT AAT GTG TTC ATC CTG CGG CTG CTG GAA AAT GCA GAT GAC Thr Leu Ala Asn Val Phe Ile Leu Arg Leu Leu Glu Asn Ala Asp Asp 385                390                395	1198
CGA GAG GCC ACC TAT ATG CTG AAG GCA TCC TCC CAG AGC GAG ATG AAG Arg Glu Ala Thr Tyr Met Leu Lys Ala Ser Ser Gln Ser Glu Met Lys 400                405                410                415	1246

- 36 -

CGC TGG ATG ACC TCA CTG GCC CCC AAC AGG AGG ACC AAG TTT GTA TCC Arg Trp Met Thr Ser Leu Ala Pro Asn Arg Arg Thr Lys Phe Val Ser 420 425 430	1294
TTC ACA TCT CGG CTG TTG GAC TGT CCC CAG GTC CAG TGT GTG CAC CCG Phe Thr Ser Arg Leu Leu Asp Cys Pro Gln Val Gln Cys Val His Pro 435 440 445	1342
TAT GTG GCC CAG CAG CCT GAT GAA CTG ACG CTG GAA CTG GCA GAT ATC Tyr Val Ala Gln Gln Pro Asp Glu Leu Thr Leu Glu Leu Ala Asp Ile 450 455 460	1390
CTG AAC ATC CTG GAG AAG ACA GAG GAT GGG TGG ATC TTT GGT GAG CGG Leu Asn Ile Leu Glu Lys Thr Glu Asp Gly Trp Ile Phe Gly Glu Arg 465 470 475	1438
CTG CAT GAC CAG GAG AGA GGC TGG TTC CCC AGT TCC ATG ACA GAG GAG Leu His Asp Gln Glu Arg Gly Trp Phe Pro Ser Ser Met Thr Glu Glu 480 485 490 495	1486
ATC CTG AAC CCC AAG ATC CGC TCC CAG AAC CTC AAG GAA TGT TTC CGG Ile Leu Asn Pro Lys Ile Arg Ser Gln Asn Leu Lys Glu Cys Phe Arg 500 505 510	1534
GTA CAT AAG ATG GAA GAC CCT CAG CGC AGC CAG AAT AAG GAC CGC AGG Val His Lys Met Glu Asp Pro Gln Arg Ser Gln Asn Lys Asp Arg Arg 515 520 525	1582
AAG CTG GGC AGC CGG AAT CGT CAA TGA ACCTCCCCAG CTCAGGCACC Lys Leu Gly Ser Arg Asn Arg Gln * 530 535	1629
TGAAGGGAAG GGTGTGGCA GGGATGGGA GCAGGCCGG CAGAGACGCC CGACAGATT 1689	
AGAGGGCCTT AGGGAAGAAT GTCAGTGCCT TCTCAGGCAG CAGGAGTGGC TTGGCCTGC 1749	
TCTGTCCCTG CCCATGCTGT GGAAGCTCTA GTGTCTGGC CACTTGTTG CTTGCACACT 1809	
GGTAAAAGC TAAGTACTTA GGCAAGTATTACACCACTCC CTTCAAGTCTC TCAGAGGTAG 1869	
AAGAAGGCAG GCATGCTCCA GAGACCTTCC GGTGACTGGA AGAGGCCAC ACAAGGGTCC 1929	
CTGGCAGCAG GCAGGTGGAA GGTAACCACT GTCAGGATCC CCTGAACCTGC ACGTGTCCCTT 1989	
CCCTACTTTG GAAGCTGTTA AGAGTCTACC AGGCACACAG ATGGCCGCC CTGCCCAGG 2049	
GAGTTTGATG AGCAGTGGTG ACCCTGCCCTG CCCGTCCCCG TGCTCTGCC AGCCTCTCTT 2109	
GCACGCCAAG CCCTGCCCTC AGCAGGCTTC CCAAAGCTTA GCTGAGGGTT CATGCCACCT 2169	

- 37 -

CTAGCTCCTT GAAGGGCTTG ATATCACTTG TGTCTCCTGG GCCCCTGATG GAGCCCAGGC	2229
GTTCGCAGA ATGAATTGGT CACTGCATCC TTTATGGTCA TGGTTTGAG AAAAGCAAAT	2289
ATCATTTTG GCTGCATTA AAGAACATC CTATATAAA AAAAAAAA AAAA	2343

## (2) INFORMATION FOR SEQ ID NO: 6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 535 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(iii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Leu	Tyr	Gln	Glu	Tyr	Arg	Asp	Lys	Ser	Thr	Leu	Gln	Glu	Ile	Glu	Thr
1										10			15		

Arg	Arg	Gln	Gln	Asp	Ala	Glu	Ile	Gln	Gly	Asn	Ser	Asp	Gly	Ser	Gln
							20		25			30			

Val	Gly	Glu	Asp	Ala	Gly	Glu	Gly	Glu							
							35		40			45			

Glu	Glu	Glu	Leu	Ala	Ser	Pro	Pro	Glu	Arg	Arg	Ala	Leu	Pro	Gln	Ile
							50		55			60			

Cys	Leu	Leu	Ser	Asn	Pro	His	Ser	Arg	Phe	Asn	Leu	Trp	Gln	Asp	Leu
							65		70			75			80

Pro	Glu	Ile	Gln	Ser	Ser	Gly	Val	Leu	Asp	Ile	Leu	Gln	Pro	Glu	Glu
							85		90			95			

Ile	Arg	Leu	Gln	Glu	Ala	Met	Phe	Glu	Leu	Val	Thr	Ser	Glu	Ala	Ser
							100		105			110			

Tyr	Tyr	Lys	Ser	Leu	Asn	Leu	Leu	Val	Ser	His	Phe	Met	Glu	Asn	Glu
							115		120			125			

Arg	Leu	Lys	Lys	Ile	Leu	His	Pro	Ser	Glu	Ala	His	Ile	Leu	Phe	Ser
							130		135			140			

Asn	Val	Leu	Asp	Val	Met	Ala	Val	Ser	Glu	Arg	Phe	Leu	Leu	Glu	Leu
							145		150			155			160

Glu	His	Arg	Met	Glu	Glu	Asn	Ile	Val	Ile	Ser	Asp	Val	Cys	Asp	Ile
							165		170			175			

- 38 -

Val Tyr Arg Tyr Ala Ala Asp His Phe Ser Val Tyr Ile Thr Tyr Val  
180 185 190

Ser Asn Gln Thr Tyr Gln Glu Arg Thr Tyr Lys Gln Leu Leu Gln Glu  
195 200 205

Lys Ala Ala Phe Arg Glu Leu Ile Ala Gln Leu Glu Leu Asp Pro Lys  
210 215 220

Cys Lys Gly Leu Pro Phe Ser Ser Phe Leu Ile Leu Pro Phe Gln Arg  
225 230 235 240

Ile Thr Arg Leu Lys Leu Leu Val Gln Asn Ile Leu Lys Arg Val Glu  
245 250 255

Glu Arg Ser Glu Arg Glu Gly Thr Ala Leu Asp Ala His Lys Glu Leu  
260 265 270

Glu Met Val Val Lys Ala Cys Asn Glu Gly Val Arg Lys Met Ser Arg  
275 280 285

Thr Glu Gln Met Ile Ser Ile Gln Lys Lys Met Glu Phe Lys Ile Lys  
290 295 300

Ser Val Pro Ile Ile Ser His Ser Arg Trp Leu Leu Lys Gln Gly Glu  
305 310 315 320

Leu Gln Gln Met Ser Gly Pro Lys Thr Ser Arg Thr Leu Arg Thr Lys  
325 330 335

Lys Leu Phe Arg Glu Ile Tyr Leu Phe Leu Phe Asn Asp Leu Leu Val  
340 345 350

Ile Cys Arg Gln Ile Pro Gly Asp Lys Tyr Gln Val Phe Asp Ser Ala  
355 360 365

Pro Arg Gly Leu Leu Arg Val Glu Glu Leu Glu Asp Gln Gly Gln Thr  
370 375 380

Leu Ala Asn Val Phe Ile Leu Arg Leu Leu Glu Asn Ala Asp Asp Arg  
385 390 395 400

Glu Ala Thr Tyr Met Leu Lys Ala Ser Ser Gln Ser Glu Met Lys Arg  
405 410 415

Trp Met Thr Ser Leu Ala Pro Asn Arg Arg Thr Lys Phe Val Ser Phe  
420 425 430

- 39 -

Thr Ser Arg Leu Leu Asp Cys Pro Gln Val Gln Cys Val His Pro Tyr  
 435 440 445  
 Val Ala Gln Gln Pro Asp Glu Leu Thr Leu Glu Leu Ala Asp Ile Leu  
 450 455 460  
 Asn Ile Leu Glu Lys Thr Glu Asp Gly Trp Ile Phe Gly Glu Arg Leu  
 465 470 475 480  
 His Asp Gln Glu Arg Gly Trp Phe Pro Ser Ser Met Thr Glu Glu Ile  
 485 490 495  
 Leu Asn Pro Lys Ile Arg Ser Gln Asn Leu Lys Glu Cys Phe Arg Val  
 500 505 510  
 His Lys Met Glu Asp Pro Gln Arg Ser Gln Asn Lys Asp Arg Arg Lys  
 515 520 525  
 Leu Gly Ser Arg Asn Arg Gln \*  
 530 535

## (2) INFORMATION FOR SEQ ID NO: 7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 803 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 3..803

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GG AGA GCT CTG CCT CAG ATC TGC CTG CTC AGT AAC CCC CAC TCA AGG	47
Arg Ala Leu Pro Gln Ile Cys Leu Leu Ser Asn Pro His Ser Arg	
1 5 10 15	
TTC AAC CTC TGG CAG GAT CTT CCC GAG ATC CGG AGC AGC GGG GTG CTT	95
Phe Asn Leu Trp Gln Asp Leu Pro Glu Ile Arg Ser Ser Gly Val Leu	
20 25 30	

-40-

GAG ATC CTA CAG CCT GAG GAG ATT AAG CTG CAG GAG GCC ATG TTC GAG Glu Ile Leu Gln Pro Glu Glu Ile Lys Leu Gln Glu Ala Met Phe Glu 35                   40                   45	143
CTG GTC ACT TCC GAG GCG TCC TAC TAC AAG AGT CTG AAC CTG CTC GTG Leu Val Thr Ser Glu Ala Ser Tyr Tyr Lys Ser Leu Asn Leu Leu Val 50                   55                   60	191
TCC CAC TTC ATG GAG AAC GAG CGG ATA AGG AAG ATC CTG CAC CCG TCC Ser His Phe Met Glu Asn Glu Arg Ile Arg Lys Ile Leu His Pro Ser 65                   70                   75	239
GAG GCG CAC ATC CTC TTC TCC AAC GTC CTG GAC GTG CTG GCT GTC AGT Glu Ala His Ile Leu Phe Ser Asn Val Leu Asp Val Leu Ala Val Ser 80                   85                   90                   95	287
GAG CGG TTC CTC CTG GAG CTG GAG CAC CGG ATG GAG GAG AAC ATC GTC Glu Arg Phe Leu Leu Glu Leu Glu His Arg Met Glu Glu Asn Ile Val 100                 105                 110	335
ATC TCT GAC GTG TGT GAC ATC GTG TAC CGT TAT GCG GCC GAC CAC TTC Ile Ser Asp Val Cys Asp Ile Val Tyr Arg Tyr Ala Ala Asp His Phe 115                 120                 125	383
TCT GTC TAC ATC ACC TAC GTC AGC AAT CAG ACC TAC CAG GAG CGG ACC Ser Val Tyr Ile Thr Tyr Val Ser Asn Gln Thr Tyr Gln Glu Arg Thr 130                 135                 140	431
TAT AAG CAG CTG CTC CAG GAG AAG GCA GCT TTC CGG GAG CTG ATC GCG Tyr Lys Gln Leu Leu Gln Glu Lys Ala Ala Phe Arg Glu Leu Ile Ala 145                 150                 155	479
CAG CTA GAG CTC GAC CCC AAG TGC AGG GGG CTG CCC TTC TCC TCC TTC Gln Leu Glu Leu Asp Pro Lys Cys Arg Gly Leu Pro Phe Ser Ser Phe 160                 165                 170                 175	527
CTC ATC CTG CCT TTC CAG AGG ATC ACA CGC CTC AAG CTG TTG GTC CAG Leu Ile Leu Pro Phe Gln Arg Ile Thr Arg Leu Lys Leu Leu Val Gln 180                 185                 190	575
AAC ATC CTG AAG AGG GTA GAA GAG AGG TCT GAG CGG GAG TGC ACT GCT Asn Ile Leu Lys Arg Val Glu Glu Arg Ser Glu Arg Glu Cys Thr Ala 195                 200                 205	623
TTG GAT GCT CAC AAG GAG CTG GAA ATG GTG GTA AAG GCA TGC AAC GAG Leu Asp Ala His Lys Glu Leu Glu Met Val Val Lys Ala Cys Asn Glu 210                 215                 220	671

-41-

GGC GTC AGG AAA ATG AGC CGC ACG GAA CAG ATG ATC AGC ATT CAG AAG Gly Val Arg Lys Met Ser Arg Thr Glu Gln Met Ile Ser Ile Gln Lys	719
225                    230                    235	
AAG ATG GAG TTC AAG ATC AAG TCG GTG CCC ATC ATC TCC CAC TCC CGC Lys Met Glu Phe Lys Ile Lys Ser Val Pro Ile Ile Ser His Ser Arg	767
240                    245                    250                    255	
TGG CTG CTG AAG CAG GGT GAG CTG CAG CAG ATG TCC Trp Leu Leu Lys Gln Gly Glu Leu Gln Gln Met Ser	803
260                    265	

## (2) INFORMATION FOR SEQ ID NO: 8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 267 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Arg Ala Leu Pro Gln Ile Cys Leu Leu Ser Asn Pro His Ser Arg Phe	
1                    5                    10                    15	
Asn Leu Trp Gln Asp Leu Pro Glu Ile Arg Ser Ser Gly Val Leu Glu	
20                    25                    30	
Ile Leu Gln Pro Glu Glu Ile Lys Leu Gln Glu Ala Met Phe Glu Leu	
35                    40                    45	
Val Thr Ser Glu Ala Ser Tyr Tyr Lys Ser Leu Asn Leu Leu Val Ser	
50                    55                    60	
His Phe Met Glu Asn Glu Arg Ile Arg Lys Ile Leu His Pro Ser Glu	
65                    70                    75                    80	
Ala His Ile Leu Phe Ser Asn Val Leu Asp Val Leu Ala Val Ser Glu	
85                    90                    95	
Arg Phe Leu Leu Glu Leu Glu His Arg Met Glu Glu Asn Ile Val Ile	
100                    105                    110	
Ser Asp Val Cys Asp Ile Val Tyr Arg Tyr Ala Ala Asp His Phe Ser	
115                    120                    125	
Val Tyr Ile Thr Tyr Val Ser Asn Gln Thr Tyr Gln Glu Arg Thr Tyr	
130                    135                    140	

- 42 -

Lys Gln Leu Leu Gln Glu Lys Ala Ala Phe Arg Glu Leu Ile Ala Gln  
145 150 155 160

Leu Glu Leu Asp Pro Lys Cys Arg Gly Leu Pro Phe Ser Ser Phe Leu  
165 170 175

Ile Leu Pro Phe Gln Arg Ile Thr Arg Leu Lys Leu Leu Val Gln Asn  
180 185 190

Ile Leu Lys Arg Val Glu Glu Arg Ser Glu Arg Glu Cys Thr Ala Leu  
195 200 205

Asp Ala His Lys Glu Leu Glu Met Val Val Lys Ala Cys Asn Glu Gly  
210 215 220

Val Arg Lys Met Ser Arg Thr Glu Gln Met Ile Ser Ile Gln Lys Lys  
225 230 235 240

Met Glu Phe Lys Ile Lys Ser Val Pro Ile Ile Ser His Ser Arg Trp  
245 250 255

Leu Leu Lys Gln Gly Glu Leu Gln Gln Met Ser  
260 265

CLAIMS

1. A polynucleotide encoding murine guanine nucleotide exchange factor (MNGEF) or a homologue thereof.
2. A polynucleotide according to claim 1 wherein said homologue is human guanine nucleotide exchange factor (NGEF).
3. A polynucleotide selected from:
  - (a) polynucleotides comprising the nucleotide sequence set out in SEQ ID No. 1, 3, 5 or 7 or the complement thereof.
  - (b) polynucleotides comprising a nucleotide sequence capable of hybridising to the nucleotide sequence set out in SEQ ID No. 1, 3, 5 or 7, or a fragment thereof.
  - (c) polynucleotides comprising a nucleotide sequence capable of hybridising to the complement of the nucleotide sequence set out in SEQ ID No. 1, 3, 5 or 7, or a fragment thereof.
  - (d) polynucleotides comprising a polynucleotide sequence which is degenerate as a result of the genetic code to the polynucleotides defined in (a), (b) or (c).
4. A polynucleotide probe which comprises a fragment of at least 15 nucleotides of a polynucleotide as defined in any one of claims 1 to 3.
5. A polypeptide in substantially isolated form which comprises the sequence set out in SEQ ID Nos. 2, 4, 6 or 8; or a polypeptide substantially homologous thereto, or a fragment of the polypeptide of SEQ ID Nos. 2, 4, 6 or 8.
6. A polynucleotide encoding a polypeptide according to claim 5.
7. A vector comprising a polynucleotide as defined in any one of claims 1 to 3 or 6.

- 44 -

8. An expression vector comprising a polynucleotide as defined in any one of claims 1 to 3 or 6, operably linked to regulatory sequences capable of directing expression of said polynucleotide in a host cell.
9. An antibody capable of binding the polypeptide of SEQ ID No. 2, 4, 6 or 8 or fragment thereof.
10. A method for detecting the presence or absence of a polynucleotide as defined in any one of claims 1 to 3 or 6 in a biological sample which comprises:
  - (a) bringing the biological sample containing DNA or RNA into contact with a probe according to claim 4 under hybridising conditions; and
  - (b) detecting any duplex formed between the probe and nucleic acid in the sample.
11. A method of detecting polypeptides as defined in claim 5 present in biological samples which comprises:
  - (a) providing an antibody according to claim 9;
  - (b) incubating a biological sample with said antibody under conditions which allow for the formation of an antibody-antigen complex; and
  - (c) determining whether antibody-antigen complex comprising said antibody is formed.
12. A polynucleotide according to any one of claims 1 to 3 or 6 for use in a method of treatment of the human or animal body.
13. A polypeptide according to claim 5 for use in a method of treatment of the human or animal body.
14. An antibody according to claim 10 for use in a method of treatment of the human or animal body.
15. A method of treating a disease or disorder of the nervous system, comprising

- 45 -

administering an effective amount of a polynucleotide as defined in any one of claims 1 to 3 or 6, to a patient.

16. A method of treating a disease or disorder of the nervous system, comprising administering an effective amount of a polypeptidetide as defined in claim 5, to a patient.
17. A method of treating a disease or disorder of the nervous system, comprising administering an effective amount of an antibody as defined in claim 10 to a patient.
18. The method of claim 15, 16 or 17 wherein said disease or disorder is a malignancy.

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 97/03302

A. CLASSIFICATION OF SUBJECT MATTER				
IPC 6	C12N15/12	C07K14/47	C07K16/18	A61K38/17
	G01N33/68			C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 C12N C07K A61K C12Q G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BOGUSKI M.S. AND MCCORMICK F. : "Proteins regulating Ras and its relatives." NATURE, vol. 366, 1993, pages 643-654, XP002057779 see the whole document, especially Table 1	1,2,4,7, 10,12, 15,18
X	HART M.J. ET AL.: "Identification of a novel guanine exchange factor for the Rho GTPase." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 271, no. 41, 11 October 1996, pages 25452-25458, XP002057776 see the whole document	1,2,4,7, 10,12, 15,18



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search

18 March 1998

Date of mailing of the international search report

27.03.98

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl  
Fax: (+31-70) 340-3016

Authorized officer

Mandl, B

## INTERNATIONAL SEARCH REPORT

International Application No.

PCT/GB 97/03302

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WHITEHEAD I.P. ET AL.: "Expression cloning of lsc, a novel oncogene with structural similarities to the Dbl family of guanine nucleotide exchange factors." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 271, no. 31, 2 August 1996, pages 18643-18650, XP002057777 see the whole document ---	1,2,4,7, 10,12, 15,18
X	CHAN A. M.-L. ET AL.: "Expression cDNA cloning of a novel oncogene with sequence similarity to regulators of small GTP-binding proteins." ONCOGENE, vol. 9, 1994, pages 1057-1063, XP002059291 cited in the application see the whole document -----	1,2,4,7, 10,12, 15,18